Information Management Advice 30 Digitisation Dilemmas

Introduction

There are substantial benefits to be gained by digitising paper records. As a result, the number of Agencies undertaking digitisation has increased dramatically in recent years.

However, if you want to introduce digitisation projects or programs you need to be aware that there are also a variety of risks and issues that Agencies should be aware of. Being aware of these risks and issues can help your Agency to build mitigation measures into program/project planning and management.

Content heading

Agencies usually conduct either:

- business process digitisation programs, where they scan incoming correspondence as it is received
- back-capture digitisation projects, where accumulated paper records are digitised

Benefits of digitisation

When programs and projects are planned and managed well and in line with best practice, digitisation can bring many benefits. Depending on the program or project, digitisation may allow you to:

- Improve the accessibility of records within the agency
- Publish records online and/or facilitate information sharing more widely
- Integrate records with existing EDRMS and business systems
- Improve efficiency by reducing reliance on hybrid systems
- Automate workflow
- Enable and encourage data reuse
- Promote the preservation of original paper records by using digital images instead
- Reduce physical storage costs by destroying paper records after digitisation.

Digitisation potential issues

Below are some examples of issues that may apply, depending on the nature of the program or project, and how issues can be managed.

Department of Education
LINC Tasmania

Tasmania
Explore the possibilities

1. That money will be wasted by poor selection of records for digitisation

Some Agencies get so excited about digitisation that they want to digitise everything! Digitisation is an expensive undertaking and should only be done when substantial and realistic benefits can be made.

Agency records will fall into different types;

- active (still being used)
- inactive (no longer used) which are for temporary or permanent retention.

For inactive temporary records there would be no point in digitising a record with a retention of 2 years as this record may already be due for destruction, and then you have to invest staff time in saving it into your records system and then completing a disposal process.

There is no point digitising short term value records or records with short retention values unless they have a high access and retrieval rate by staff and measurable business efficiencies can be gained.

It is also worth noting that some records do not digitise well e.g. fragile records, records with faint inks etc. Others, such as large maps and plans, will require specialist (and often more expensive) digitisation techniques and equipment.

Solution: Selection decisions for digitisation programs or projects should be carefully defined in planning. There should be defined business efficiency benefits as a result of digitisation projects.

2. That the main business driver for digitisation should be business efficiency gains

TAHO recommends that you look at the types of records you hold, their value (temporary or permanent), and their business use - how often are they called upon and used. Discuss with the records staff and users - which records are used frequently (daily) and where efficiency gains can be made by digitising them for staff. The main drivers of digitisation projects are:

- if you work in a distributed network of officers (all around Tasmania) and many staff in different locations require assess to the same documents
- if you need to access information rapidly i.e. there is a customer at the desk and you need access to their file straight away
- if you refer to the same information regularly i.e. it's retrieved daily
- and for disaster recovery and business continuity reasons.

Solution: If these business drivers do not exist then you need to consider if you will make the efficiency gains required to effectively justify the cost of mass digitisation to your executive management.

3. That the complexities of digitisation will be underestimated

Digitisation is inherently complex. It is not enough just to buy the software and hardware and start scanning, or sign a contract with a service provider and send them paper records. If care is not taken the results could be disastrous, the digital images unusable and public money wasted.

Solution: All digitisation activity needs to be carefully planned and managed. Research needs to be conducted and decisions need to be made and documented about a variety of matters including:

- what records should be digitised and why
- what benchmarks are needed

what technical standards to use (see *Guideline 19 Digital Preservation Formats*)

- what metadata is required, and
- what quality assurance measures to apply.

Staff members need to be trained and supported to perform their allocated roles.

If contracting-out digitisation, agencies need to ensure that controls are communicated to service providers and built into contracts. The performance of the contractor must be monitored and the paper and digital records protected at all times. For more information about responsibilities when contracting-out. (see *State Records Guideline 10 Outsourcing of government business: recordkeeping issues*).

4. That costs will be underestimated

Digitisation involves a range of upfront and ongoing costs

- Upfront costs can include the purchase of hardware and software, training and support of staff, staff time to define benchmarks, contract set-up etc. In some cases costs might include conservation treatment for fragile paper records prior to digitisation.
- Ongoing costs include staff time to digitise and perform quality assurance, the cost of maintaining digital images (and possibly original paper records as well) over time, managing variables for example, image resolution and output file format which will affect the quality of the resultant digitised image and the size of the file these need to be reviewed and updated regularly to ensure they are fit for purpose and the maintenance and update of systems.

In particular, the ongoing costs (and complexity) of preserving digital images in place of original paper records over long periods of time should not be underestimated. This will require a significant investment of funds over a long period of time which could make some projects unviable.

If outsourcing, some upfront costs can be reduced, but the agency will incur other costs from the service provider, or additional work that they may need to complete internally.

Solution: As part of planning, agencies need to clearly define both upfront and ongoing costs, including costs for maintaining some digital images in the long term, and ensure that they have adequate resources over the lifetime of the program or project. If they cannot guarantee resources, project aims and selection decisions may have to be reviewed.

5. Migration of digitised documents into your records system needs to be considered

Processes for migrating digitised documents into your records system needs to be considered. If this is going to be done manually by your records staff it will be quite an additional workload as all of these documents will need to be saved into the system on top of their normal work. If your IT staff are going to bulk load them into the system then you will need to consider the IT process and limitations around this and define these for the vendor, i.e. what are the IT section's required file formats and loading process requirements? Can the vendor deliver the documents in the formats your IT section requires and that are appropriate for long term preservation?

Solution: As part of the agency's planning process, define how the records and going to be migrated into your records systems and the technical and business process requirements.

6. That the digitised records are in appropriate formats

TAHO recommends that permanent records be digitised in certain formats, for example documents in PDF/A format, to ensure their longevity over time. See *State Records Guideline 19 Digital Recordkeeping Formats*.

7. That multiple documents will be scanned into one PDF

Some agencies manage information on case files, for example Property files, Personnel files or Patient Files. These contain both temporary and permanent value records. Many agencies are now looking to digitise the content of case files. As a cost saving measure, vendors will often offer to digitise the entire file into one PDF with OCR capabilities. This means that the content of the document can be searched and retrieved but essentially many different records are a part of one single PDF. Whilst this can be a cost saving measure, it will not be practical to manage in the long term.

In order for the agency to archive permanent records and destroy temporary records, the agency will need to split the PDF into separate PDFs. This is possible by purchasing a full version of Adobe Acrobat (not just Acrobat reader). Then a business process will need to be put in place to split the PDF into separate documents saving the permanent value records as separate documents into the agency's EDRMS. This can be quite a labour intensive process and may require several staff members depending upon the size of the initial digitisation project. If the agency does not do this, then they will not be permitted to destroy the paper records because the digitised copy would not be considered fit for purpose. TAHO will only accept the permanent value records so the agency will need to implement this business process in order to transfer records.

Solution: Consider the ramifications of implementing these types of business process in the planning stages and build in contingencies to address them.

8. That metadata with the scanned images will not be sufficient

The agency will need to define for the vendor the standard for titling the documents so that your staff have enough information about the file to effectively migrate it into your system without having to open and look at every single document.

The migration on large batches of digitised records into your EDRMS can often be automated by your IT people, however to make this work they will need to define for you the metadata requirements for their scripts (i.e. how they are going to match particular records with particular files).

Solution: Define metadata requirements in the planning stages of your project and discuss your requirements with the vendor and ensure it is detailed in the contract.

9. That there are adequate quality assurance and review processes for digitised records

It will be necessary for the agency to specify in the contract or agreement the level of quality review or assurance (QA) of the scanned images they require the vendor to undertake.

The vendor should undertake to QA at least 80% of your digitised content, and for permanent records TAHO recommend QA be 100% and that you include an internal QA process as a part of your project. This is where the cost is for vendors, so if the level of QA you are being offered is low and the cost is low, then that means you risk receiving batches of digitised content that may have well over a 50% error rate. Errors can occur easily as soon as you start putting large batches of documents through industrial scanners. If your contract does not specify the required level of QA, then you will have no recourse to get it fixed and it will be wasted effort.

10. That original Building files may be accidentally destroyed

Paper copies of building files can be destroyed once they are scanned as long as they meet the conditions set out in DA2159, and if they do not include the following:

- Files relating to properties on the councils heritage register
- Files relating to buildings identified by the council as having significant architectural or cultural value
- Files containing building plans with an intrinsic artistic value

These paper files cannot be destroyed.

11. That Historical pre 1960s records will be treated as Scheduled

Historical records or records created prior to 1960 are not covered by any Disposal Schedule and consequently cannot be destroyed without approval from the State Archivist. As there are so few historical records surviving, TAHO would appraise any pre 1960s records to see if we would be interested in accepting them into TAHO.

12. That original paper records, including State archives, will be destroyed without authorisation

Any agencies undertaking digitisation in order to destroy original paper records need to be aware that not all records are authorised for destruction after digitisation. If a reduction in storage costs is one of the promised benefits in their business case, this may not always be deliverable.

Where original paper records can be destroyed, the digital images will become the primary records and need to be managed accordingly and kept for the full retention periods specified in retention and disposal authorities.

Solution: Agencies should clearly understand disposal requirements for the original paper records selected for digitisation before embarking on a digitisation program or project. They should also build in staff training and quality assurance measures to ensure that no unauthorised disposal takes place. Any authorised destruction should be secure and in line with the agency's Records Policy.

13. That original paper records will be damaged during the digitisation process

Back-capture digitisation projects require that original paper records be retrieved from storage, handled and sometimes unbound. On occasions pages may need to be removed (e.g. in order to apply different scanner settings). In the case of outsourcing, original paper records will need to be transported to and from service providers.

There is a risk that the paper records will be damaged in this process or returned to storage in the incorrect order. This may create problems and compromise their ability to function as evidence, especially for records of high risk or high value that are being retained after digitisation.

Solution: Agencies need to ensure that original paper records are handled carefully and put back in their original order. Staff members need to be trained in handling, and supervised where necessary and quality assurance measures built in to ensure records are not damaged.

Important note: If records are required as State archives, advice from TAHO should be sought before digitisation commences.

14. That digitisation will not produce a full and accurate representation of the original paper records

The aim of digitisation is to create digital images of the original paper records. The quality of these images need to be fit for purpose. In some cases it may be appropriate to settle for smaller files of lesser quality e.g. for the delivery of images online.

However, there are occasions where the quality of the digital images cannot be compromised, for example, where the original paper records are to be destroyed and the digital records are to replace them as evidence of the business activity (and could potentially be required in court).

Solution: Agencies need to identify what the essential characteristics of the records are that must be retained, and create quality benchmarks to ensure that these remain intact. The definition of technical and metadata standards, along with staff training and quality assurance checks (either of images or samples of images) should help to ensure that image quality is fit for purpose. Procedures and standards need to be developed and documented as these will help to demonstrate that digital images are trusted copies of the originals.

15. That digital images are not stored or protected appropriately

If the original paper records are destroyed the digital images become the official records. Digital images must be stored securely, remain tamper-proof and unaltered, and be kept for the retention period defined for the original paper records. Poor storage or protection of the digital images may result in the loss of essential evidence, which can have far-reaching consequences for the agency.

Even if the original paper records are retained, there are risks to the agency if digital images are not subject to the same controls as the original paper records. For example, access and security regimes for paper records with sensitive content should equally be applied to the digital images.

Solution: Agencies should confer with their records management staff to ensure that digital images are stored and controlled appropriately.

Ideally, digital images should be captured into a recordkeeping system, where they can be accessed only by authorised users, secured from unauthorised alteration or deletion, in context with related records, protected from disaster and kept for their required retention periods.

If the agency does not have a recordkeeping system, digital images need to be kept on a secure drive with stringent security, access controls and backup until the agency can move them into a recordkeeping system.

Note: Master digital images should not be stored on removable media due to their inherent risks.

16. That duplicates are not managed appropriately

When original paper records are retained, the digital (master) image is already a duplicate. Agencies may also potentially require multiple digital copies at different levels of quality, depending on their purpose.

If these are inefficiently managed, it will be difficult to ascertain which version is the "official" record and to ensure that it survives for as long as it is needed. The existence of many derivative images may also add to storage and management costs and cause problems in relation to access and security.

Solution: Agencies should confer with their records management staff to ensure that digital images are managed appropriately. They will also need to adopt a well-defined metadata schema with relationships between records, to ensure that it is very clear whether the original paper record exists, which is the "master" image and which are derivatives and for what purpose they are to be used.

17. That digital images will not survive and remain accessible for as long as they are required

Digital images are inherently unstable. Technological obsolescence and media degradation can render them unreadable in relatively short time frames. If digital images are replacing original paper records, and are required as evidence for long periods of time, the agency will need to consider how to ensure they survive for as long as they are required.

Solution: Agencies should confer with their records management staff to ensure that digital records are managed as part of a records management program and that their longevity is planned for. For example, the agency may be required to plan for, test and implement data and systems migration on an average of 5-7 years in order to maintain information accessibility and integrity. Funds for ongoing management should be defined in business cases and made available when required. Technological and storage decisions will impact on the longevity of images and should be made with care.

In Summary

Current thinking around digitisation is that it should be undertaken "on demand" basis for inactive files. This is because the cost of quality digitisation is so high and the management effort by the business to gain the efficiencies required to make it worthwhile, are considerable.

Digitisation projects require planning. If the issues discussed here are not considered, then a digitisation project can create a huge mess, not unlike a room full of records that nobody knows what they are, or a folder on a network drive with thousands of documents that staff do not have time to process.

The above issues and any other risks need to be identified and should be considered as part of the planning and management of digitisation programs or projects. If mitigation strategies are built in, agencies can enjoy the benefits digitisation can offer, while avoiding the pitfalls.

Guidelines

TAHO has developed a Digitisation toolkit, both for business process digitisation programs and back-capture digitisation projects, which addresses how to manage projects, optimise benefits and reduce risks. This toolkit is available via the GISU website.

Please feel free to provide us with feedback by email at gisu@education.tas.gov.au if:

- you want to discuss any particular digitisation problems you are encountering
- you believe you have established a best practice digitisation program or project
- you have found innovative solutions to common digitisation problems
- you would like to share your business case for digitisation.

Such information will help us to ensure that the guidelines are beneficial for all agencies undertaking digitisation.

Further Advice

For more detailed advice please contact:

Government Information Strategy Unit
Tasmanian Archive and Heritage Office
91 Murray Street
HOBART TASMANIA 7000
Telephone: 03 6165 5581
Email: gisu@education.tas.gov.au

Information Security Classification

This document has been security classified using the Tasmanian Government Information Security classification standard as PUBLIC and will be managed according to the requirements of the Tasmanian Government Information Security Policy.

Document Development History
Build Status

Version	Date	Author	Reason	Sections
2.0	March 2015	Christine Woods	Template	All
1.0	28-02-2013	Allegra Huxtable	Initial Release	All

Amendments in this Release

Section Title	Section Number	Amendment Summary
All	All	Document imported into new template

Issued: February 2013

Ross Latham
State Archivist